

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28

2
3

4
5

6

7
8
9
10
11
12

13

14

14
15
16
17

18
19

19
20
21
22
23
24
25
26
27

1 November 2002 by Springer-Verlag, Berlin and Heidelberg, Germany;
2 and C. Chung, et al., "DEMIDS: A Misuse Detection System for
3 Database Systems", Department of Computer Science, University of
4 California at Davis, Davis, California, October 1, 1999.

5
6 A common flaw in database intrusion detection systems of the
7 prior art is that such systems fail to protect the database
8 against insider attempts to steal large amounts of data using
9 legitimate business processes. For example, such a system may
10 allow a given service representative to access fields and tables
11 within the database containing customer credit card information.
12 Normally, a representative might access 5 to 10 accounts per hour
13 in order to service customers. That is fine until the customer
14 service representative decides to launch an insider attack on the
15 database, procuring large amounts of consumer credit card
16 information, which he then uses for nefarious purposes. The
17 present invention is designed to protect against that and other
18 attacks.
19

20 Disclosure of Invention

21 Methods, apparatus, and computer-readable media for
22 protecting computer code (1) from malicious retrievers (3). A
23 method embodiment of the present invention comprises the steps of
24 generating (22) retrieval information characteristic of data sent
25 to a retriever (3) by the computer code (1) in response to a
26 retrieval command (5) issued by the retriever (3); accessing at
27
28

1 least one rule (6) using at least some of said retrieval
2 information as an input to said at least one rule (6); and, when
3 said at least one rule (6) informs that the retrieval is not
4 acceptable, flagging (28) the retrieval command (5) as
5 suspicious.
6

7 Brief Description of the Drawings

8 These and other more detailed and specific objects and
9 features of the present invention are more fully disclosed in the
10 following specification, reference being had to the accompanying
11 drawings, in which:

12 Figure 1 is a block diagram illustrating embodiments of the
13 present invention.
14

15 Figure 2 is a flow diagram illustrating an operational phase
16 of the present invention.

17 Figure 3 is a flow diagram illustrating a training phase of
18 the present invention.

19 Figure 4 is a flow diagram illustrating a system
20 administrator phase of the present invention.

21 Figure 5 is a diagram illustrating typical contents within
22 state table 18 of the present invention.
23

24 Figure 6 is a diagram illustrating typical contents within
25 rule table 6 of the present invention.
26
27
28

Detailed Description of the Preferred Embodiments

This invention has applicability to any code intrusion detection system, i.e., any system in which computer code 1 is susceptible to being attacked by commands 5 which may be malicious, due to malicious intent on the part of the user 3 who issues the command 5. As used herein, "user" can refer to a client computer 3 and/or to a human who has control of computer 3. As illustrated in Figure 1, there can be a plurality N of users 3, where N is any positive integer. "User" is sometimes referred to herein as "retriever".

Most of the following description illustrates the special case where the computer code 1 is a database 1. Database 1 can be any type of database, such as a relational database or a flat file. When database 1 is a relational database, commands 5 are typically written in a SQL language. As used herein, "SQL" is taken in the broad sense to mean the original language known as SQL (Structured Query Language), any derivative thereof, or any structured query language used for accessing a relational database. In the case where computer code 1 is not a relational database, the commands can be written in another language, such as XML. Database 1 may have associated therewith an internal audit table 11 and/or an external database log file 12 for storing audit and/or ancillary information pertaining to database 1. Database 1 is typically packaged within a dedicated computer

1 known as a database server 2, which may also contain database
2 communications module 15 and other modules not illustrated.

3 Computer code intrusion detection system (IDS) 19 (and its
4 special case, database intrusion detection system 19) encompasses
5 modules 4, 6-9, 13, 17, and 18. Modules 1, 4, 6-9, 11-13, 15,
6 17, and 18 can be implemented in software, firmware, hardware, or
7 any combination thereof, and are typically implemented in
8 software. Figure 1 illustrates the case where modules 4, 6-9,
9 13, 17, and 18 are stand-alone modules separate from database
10 server 2. However, these modules could just as well be
11 incorporated within database server 2, e.g., they could be
12 incorporated within database communications module 15. Thus,
13 intrusion detection system 19 could be published by a third party
14 as a standalone package on any type of computer-readable medium,
15 or bundled by the manufacturer of the database 1 with module 15.
16 The purpose of intrusion detection system 19 is to protect
17 computer code 1 from users 3 that have nefarious intent. For
18 example, such users may desire to steal (possibly large amounts
19 of) credit card information from database 1.
20
21

22 One method embodiment of the present invention comprises
23 three phases: a training phase, a system administrator phase,
24 and an operational phase. Figure 2 illustrates the operational
25 phase of the present invention. At optional step 20, computation
26 module 7 extracts an input vector from a retrieval command 5,
27
28

1 using any technique of real-time auditing and/or in-line
2 interception described below in conjunction with step 32. The
3 extraction is typically done in real time or quasi-real-time. As
4 used herein, "real time" means "during a short time interval
5 surrounding the event". Thus, observing a command 5 in real time
6 means that the command 5 is observed during a short time interval
7 surrounding the instant that the command 5 enters the database 1.

9 A retrieval command 5 is any command by which a retrieving
10 user 3 seeks to retrieve information from the database 1. The
11 input vector characterizes the retrieval command 5 and comprises
12 at least one parameter from the group of parameters comprising:
13 canonicalized commands; the dates and times at which the commands
14 5 access the computer code 1; logins (user IDs, passwords, catch
15 phrases, etc.) of users 3 issuing the commands 5; the identities
16 of users 3 issuing the commands 5; the departments of the
17 enterprise in which the users 3 work, or other groups to which
18 the users 3 belong; the applications (i.e., software programs or
19 types of software programs) that issue the commands 5; the IP
20 addresses of the issuing computers 3; identities of users 3
21 accessing a given field or fields within the computer code 1; the
22 times of day that a given user 3 accesses a given field or fields
23 within the computer code 1; the fields or combination of fields
24 being accessed by given commands 5; and tables or combinations of
25 tables within the computer code 1 accessed by the commands.
26
27
28

1 A canonicalized command is a command 5 stripped of its
2 literal field data. Literal field data is defined as a specific
3 value of a parameter. Thus, for example, let us assume that the
4 command 5 is:

5 SELECT NAME FROM PATIENTS WHERE NAME LIKE 'FRANK' AND AGE > 25

6 In this case, the literal field data is "FRANK" and "25".
7
8 Thus, a canonicalized form of the command 5 is:

9 SELECT NAME FROM PATIENTS WHERE NAME LIKE * AND AGE > *

10 Literal fields can include literal numbers (plain numbers),
11 dates, times, strings, and potentially named ordinal values
12 (symbolic words used to represent numbers, e.g., "January"
13 represents the first month, "Finance" represents department 54,
14 etc.).

15 In one embodiment, a retrieval command 5 is subjected to
16 step 20 only if the fields mentioned in the command 5 appear on a
17 preselected list of fields deemed to be important, e.g., credit
18 card and password fields. In other embodiments, the operational
19 phase is performed without the need to extract an input vector,
20 and thus step 20 is not performed at all.

21 At step 21, the retrieval command 5 is forwarded to the
22 database 1 for processing. When the database 1 finishes
23 processing the retrieval command 5, it normally sends back to
24 user 3 the requested data in the form of rows plus columns and/or
25 tables. A single row of data may contain a credit card number,
26
27
28

1 expiration date, and customer name, i.e., three columns worth of
2 data. A second row of data then would contain a second credit
3 card number, a second expiration date, and a second customer
4 name.

5
6 At step 22, computation module 7 observes this response by
7 database 1 (using any technique of real-time auditing and/or in-
8 line interception described below in conjunction with step 32);
9 and generates retrieval information therefrom. This retrieval
10 information is optionally stored in state table 18, potentially
11 along with one or more pieces of information from the input
12 vector (e.g., to maintain data such as "users of the SUPPORT
13 group retrieved an average of 10 customer records per hour").
14 State table 18 can maintain statistics on client 3 access to
15 particular fields, associating the client 3 with the types of
16 data that the client 3 is accessing. Clients 3 can be identified
17 by user-ID ("Carey"), group-membership ("Average statistics for
18 all members of the FINANCE group"), group-ID ("FINANCE group"),
19 as well as potentially source IP address, machine name
20 identification, client application, or other combinations of zero
21 or more elements of the input vector. State table 18 stores a
22 set of statistics associated with one or more of these client 3
23 identifiers. State table 18 may also group its data based on
24 other attributes in the input vector, including the set of
25 referenced fields, etc. (see point 8 below). For example:
26
27
28

1 CAREY's statistics:

- 2 1. has downloaded 2000 credit card rows total
- 3 2. downloads credit card rows at a rate of 10 per hour
- 4 during business hours
- 5 3. downloads credit card rows at a rate of 3 per hour during
- 6 off hours
- 7 4. has downloaded 1500 password rows total
- 8 5. downloads password rows at a rate of 10 per hour during
- 9 business hours
- 10 6. downloads password rows at a rate of 3 per hour during
- 11 off hours
- 12 7. downloads password rows at an average rate of 3 per
- 13 request
- 14 8. For commands that attempt to access fields {USER,
- 15 PASSWORD, SSN}, the average number of retrieved rows is
- 16 1.
- 17 9. etc...

20 FINANCE's average user statistics:

- 21 1. has downloaded 23000 credit card rows total.
- 22 2. average finance user downloads credit card rows at a rate
- 23 of 7 per hour during business hours
- 24 3. downloads credit card rows at a rate of 1 per hour during
- 25 off hours
- 26 4. etc...

1 statistics for computer at IP Address 1.2.3.4:

2 etc.

3 etc.

4 etc.

5 The statistics can be maintained for only those fields
6 deemed critical by the database administrator 10, or for all
7 fields accessed. Clearly, many types of statistics can be
8 maintained, including:
9

- 10 1. average number of row retrievals per given time unit
11 (minutes, hours, seconds)
- 12 2. standard deviation of row retrievals per given time
13 unit
- 14 3. average number of columns retrieved per time unit, etc.

15 Typical contents of a state table 18 having three entries
16 are illustrated in Figure 5. In the first entry, an input vector
17 was not calculated (at step 20), because here the operational
18 phase is operating on a command by command basis. Thus, there is
19 no need to track any identifying information for a particular
20 command 5, because it is the present command 5 that is being
21 processed.
22

23 "Retrieval information" consists of two components: one or
24 more retrieval vectors, and statistical information. As used
25 herein, "retrieval vector" comprises at least one of the
26 following: the number of rows retrieved; the number of columns
27

28

1 retrieved; the number of tables retrieved; an identification of
2 the columns that were retrieved; and an identification of the
3 tables that were retrieved. Thus, in the present example of
4 entry 1, the retrieval vector can be represented as [5 rows; 3
5 columns; columns A, J, and K]. As used herein, "statistical
6 information" means any statistics that can be generated from the
7 retrieval, either in conjunction with data stored in state table
8 18, or on its own. Thus, "statistical information" can comprise
9 one or more of the following statistics: the rate of retrieving
10 rows; the rate of retrieving columns; the rate of retrieving
11 tables; the average number of rows retrieved per retrieval
12 command 5 for a given input vector (or subset of an input
13 vector); the average number of columns retrieved per retrieval
14 command 5 for a given input vector; the average number of tables
15 retrieved per retrieval command 5 for a given input vector; the
16 percentage of retrieval commands 5 for which a given column is
17 accessed; the percentage of retrieval commands 5 for which a
18 given table is accessed; the percentage of retrieval commands 5
19 for which a given combination of columns is accessed; and the
20 percentage of retrieval commands 5 for which a given combination
21 of tables is accessed.

22
23
24
25 Note that some of these statistics are compilable across
26 many commands 5, and some are compilable within a single command
27 5. In the present example of entry 1 in Figure 5, there are two
28

1 pieces of statistical information that have been generated by
2 computation module 7 as a result of this particular command 5
3 accessing this particular database 1: S, the number of rows per
4 second that are retrieved; and D, the number of columns per
5 second that are retrieved. In this example, S=2000 rows per
6 second and D=2300 columns per second.
7

8 At step 23, computation module 7 uses retrieval information
9 to access at least one rule 6 pertaining to retrievals. The
10 rules 6 can define acceptable and/or unacceptable retrievals, and
11 can be stored in any manner known to one of ordinary skill in the
12 art. In one embodiment, at least one rule 6 comprises a pre-
13 established table containing rules for acceptable and/or
14 unacceptable retrievals as illustrated in Figure 6. In the
15 illustrated example, rule table 6 has four entries. In the first
16 entry, there is no input vector, since the corresponding rule is
17 independent of any particular input vector. (It may be said that
18 the input vector is wildcarded.) This emphasizes the fact that
19 it is not necessary for table 6 to be accessed (indexed) by an
20 input vector. In this example, the cognizant rule, rule 5,
21 states: "no more than 1000 rows per second can ever be retrieved
22 by anybody".
23
24

25 At step 26, computation module 7 determines whether table 6
26 indicates that the retrieval is acceptable or unacceptable. The
27 matching of the retrieval information from table 18 to the rule
28

1 in table 6 can be performed by any technique known to those of
2 ordinary skill in the art. If table 6 indicates that the
3 retrieval is acceptable, the retrieval is allowed to proceed at
4 step 27, i.e., the requested data is sent to the requesting user
5 3.
6

7 If, on the other hand, the retrieval information from table
8 18 does not satisfy the corresponding rule in table 6, module 8
9 flags the current command 5 as being suspicious at step 28. Then
10 a post-flagging protocol is performed by module 9 at step 29. In
11 the illustrated example, the retrieval information "S=2000 rows
12 per second" violates the rule "no more than 1000 rows per second
13 can ever be retrieved by anybody". Thus, steps 28 and 29 are
14 executed.
15

16 Execution of the post-flagging protocol at step 29 entails
17 execution of at least one of the following steps: an alert is
18 sent to the system administrator 10; an audit log is updated; the
19 command 5 is not allowed to access the computer code 1; the
20 command 5 is allowed to access the computer code 1, but the
21 access is limited in some way (for example, the amount of data
22 sent back to user 3 is limited); the command 5 is augmented,
23 e.g., investigational code is inserted into the command 5 to
24 provoke an audit trail; the user 3 sending the command 5 is
25 investigated. The latter investigation can be performed by
26 computer means (e.g., sending out a digital trace to determine
27
28

1 the identity of the user 3) and/or by off-line means (sending a
2 human private investigator to spy on user 3).

3 The above example illustrates an embodiment in which table 6
4 is accessed by retrieval information but not by an input vector.
5 In other embodiments, an input vector (or more than one input
6 vector, as long as the input vectors are from the same command
7 5), in addition to retrieval information, is used to access table
8 6. For example, consider the second entry illustrated in Figure
9 6. The four rules set forth in said entry 2 are associated with
10 a particular input vector $L_1F_1A_1$. These rules, which are more
11 fully described below in conjunction with the training phase, are
12 valid only with respect to specific input vector $L_1F_1A_1$.

13 The above examples illustrate the case where the operational
14 phase is performed on a command by command basis. In other
15 embodiments, the retrieval information can be compiled on other
16 bases, for example, with respect to all commands 5 that are
17 executed during a given time period that defines the operational
18 phase, or for the duration of a login by a user 3 to the database
19 1. This is illustrated in entry 2 of Figure 5, where the
20 retrieval information is presented without regard to input
21 vector. In this example, the retrieval information that has been
22 compiled in table 18 is the statistic "the rate of retrieving
23 rows was 2000 rows/second across all commands 5". In this
24 example, at step 26, rule 5 from table 6 remains violated, this
25
26
27
28

1 time for the operational phase taken as a whole. Thus, at step
2 28, the entire operational phase is flagged as being suspicious,
3 and the post-flagging protocol 29 performed at step 29 is
4 tailored accordingly.

5 At step 26, all of the retrieval information in state table
6 18 can be matched against all of the rules in table 6, or just a
7 subset of the retrieval information and/or a subset of the rules
8 can be used for matching.

10 An example of an embodiment where table 6 is accessed by two
11 input vectors within the same command 5, as well as by retrieval
12 information from table 18, is illustrated in entries 3 and 4 of
13 Figure 6. Entry 3 gives the rule (rule 6) that for input vector
14 L_1 , "no retrievals are allowed between 6 p.m. and midnight unless
15 rule 7 is satisfied". Let us assume that L is the log-in of the
16 user 3 issuing the command 5; L_1 is "Abacus 34"; and retrieval
17 information stored in table 18 for this command 5 specifies that
18 the command 5 was issued at 8 p.m. Then at step 26, computation
19 module 7 determines that rule 6 is violated, unless rule 7 is
20 satisfied. Thus, table 6 must also be accessed by the second
21 input vector, F_1 . Let us assume that F is the field being
22 queried by the command 5 and F_1 is the credit card number. Then,
23 computation module 7 looks to table 18 to determine whether the
24 credit card number field is retrieved at a rate D less than 10
25 per minute by that particular command 5.

1 The contents of table 6 are generated during an optional
2 training phase, and/or are force fed into table 6 by system
3 administrator 10, and/or are provided by a security or other
4 vendor. A typical training phase is illustrated in Figure 3, and
5 is initiated at step 31. This is done by system administrator 10
6 flipping a switch (which may be located, for example, on database
7 server 2 or on training module 4); by means of a preselected
8 event occurring (e.g., the first of each month or the addition of
9 a new table within database 1); or by any other means known to
10 one of ordinary skill in the art for starting a computer system.

12 At step 32, training module 4 observes retrieval commands 5
13 that users 3 send to database 1. This observation may be done in
14 real time. There are two major ways in which the observing step
15 32 can be performed: real-time auditing and in-line interception.
16 Real-time auditing is typically used in cases where database 1
17 has an auditing feature. The auditing information may be placed
18 into an audit table 11 internal to database 1 or into an external
19 database log file 12. In real-time auditing, training module 4
20 instructs the database 1 to generate a stream of events every
21 time a command 5 enters database 1. The stream can include such
22 items as the text of the command 5, a date/time stamp,
23 information pertaining to the user 3 that issued the command 5,
24 the IP (Internet Protocol) address of the issuing computer 3, the
25 application that issued the command 5, etc.

1 The stream can appear in string or binary form, and can be
2 extracted using a number of different techniques, depending upon
3 the implementation of the IDS 19, including APIs (Application
4 Programming Interfaces) that access the computer code 1. One
5 example is to use ODBC (Open DataBase Connectivity), a set of C
6 language API's that allows one to examine or modify data within
7 database 1. If the Java programming language is used, JDBC (Java
8 DataBase Connectivity) can be used instead. Another way of
9 extracting the needed information from database 1 is to use code
10 injection or patching to inject logic into one or more modules
11 1,15 within database server 2, to transfer control to training
12 module 4. In another embodiment, called "direct database
13 integration", the database 1 vendor, who has access to the
14 commands 5 in conjunction with the normal operation of the
15 database 5, makes the commands 5 available to intrusion detection
16 system 19. In yet another embodiment, in cases where database 1
17 supports it, external database log file 12 may be examined
18 without the need to resort to special software. Once a retrieval
19 command 5 has been processed by training module 4, the command 5
20 can optionally be expunged from any table or log file it is
21 stored in, to make room for subsequent commands 5.

22 In in-line interception, at least one of a proxy, firewall,
23 or sniffer 13 is interposed between database 1 and users 3 (see
24 Fig. 1). The proxy, firewall, and/or sniffer 13 examines packets

1 of information emanating from users 3 and extracts the relevant
2 information therefrom. Proxy, firewall, and/or sniffer 13 may
3 need to decrypt the communications emanating from users 3 if
4 these communications are encrypted.

5 After a command 5 has been captured in step 32, at step 33
6 training module 4 observes (extracts) the response of database 1
7 to the command 5, and updates (augments) state table 18. Step 33
8 can be performed in real time, i.e., state table 18 can be
9 updated response-by-response. The responses of the database 1
10 can be extracted using any of the techniques of real-time
11 auditing and/or in-line interception that are described above in
12 conjunction with step 32. Similarly, previously described steps
13 20 and 22 can be performed using any of the above-described
14 techniques of real-time auditing and/or in-line interception,
15 with computation module 7 rather than training module 4 doing the
16 extraction and generation, respectively.

17 The operation of step 33 is illustrated in entry 3 of Figure
18 5. The retrieval information comprises, for the illustrated
19 input vector $L_1F_1A_1$, two retrieval vectors plus statistical
20 information comprising the number of occurrences of each of the
21 retrieval vectors, plus S and D.

22 Let us assume that L is the parameter "log-in of the user 3
23 that issued the command 5". The log-in can be some preselected
24 combination of user ID, password, and answer to a challenge

1 phrase (e.g., "what is your mother's maiden name?"). In this
2 example, L_1 is "Abacus34". F is the field being queried by the
3 command 5. F_1 is "credit card number". A is the application
4 that issued the command 5 or the IP address of the requesting
5 computer 5. A_1 is "Siebel CRM Deluxe Version 22". Let us
6 further assume that during the entirety of the training phase,
7 the only responses generated by database 1 to commands 5
8 associated with $L_1F_1A_1$ are a plurality of responses having five
9 rows and three columns (retrieval vector 1), and a plurality of
10 responses having seven rows and two columns (retrieval vector 2).
11 Let us further assume that retrieval vector 1 has occurred 963
12 times, and retrieval vector 2 has occurred 51 times. Thus, the
13 augmentation of state table 18 performed in step 33 for a given
14 command 5 may simply entail incrementing the number of
15 occurrences from 962 to 963, and recalculating S and D . In the
16 illustrated example, the rate S of rows returned by database 1
17 for this input vector is 1.1 row per second, and the rate D at
18 which database 1 returns columns for this input vector is 2.3
19 columns per second.

20
21
22 Note that not all of the possible parameters have to be
23 covered in the input vector that is the subject of the training.
24 In this case, just three parameters (out of the many more
25 possible parameters) are so covered (the set of parameters to use
26 may be specified by an administrator 10).
27
28

1 Steps 32 and 33 are repeated for each command 5 that is
2 processed during the training phase.

3 The training phase is ended, at step 34, by any one of a
4 number of means. For example, system administrator 10 can flip a
5 switch on database server 2 or training module 4. Alternatively,
6 the training phase may end by a statistical technique, e.g.,
7 training module 4 monitors the occurrence or frequency of new
8 commonly occurring retrieval vectors. Alternatively, the
9 training phase may end by the occurrence of a preselected elapsed
10 or absolute time, or by any other means known to one of ordinary
11 skill in the art. As with all of the preselected parameters in
12 this patent application, the preselected parameters mentioned in
13 this paragraph may be stored in parameters storage area 17.
14

15 At step 35, module 7 converts the retrieval information
16 stored in state table 18 into rules for acceptable and/or
17 unacceptable retrievals within table 6, using preselected set of
18 parameters 17. The administrator 10 may be asked to review
19 and/or augment these rules. Entry 2 of Figure 6 corresponds to
20 entry 3 of Figure 5. There are four rules illustrated for said
21 entry. It can be seen that Rule 1 was derived from the retrieval
22 information in Figure 5 by first concluding that the 963
23 occurrences of five rows and three columns was greater than a
24 preselected threshold value (e.g., 50) to warrant inclusion in
25 table 6. Then, a preselected margin (in this case, one) in
26
27
28

1 either direction was applied around the observed numbers of rows
2 and columns to generate the rule. The "AND" following the
3 semicolon in rule 1 is a Boolean AND, i.e., both the criterion
4 "between 4 and 6 rows" and the criterion "between 2 and 4
5 columns" must be satisfied in order for the retrieval to be
6 deemed acceptable at step 26. There may also be Boolean logic
7 underlying the combination of the rules. For example, in order
8 for module 7 to conclude in step 26 that a retrieval is
9 acceptable, it might have been preselected that either Rule 1 AND
10 Rule 3 AND Rule 4 must be satisfied; OR Rule 2 AND Rule 3 AND
11 Rule 4 must be satisfied in order for the retrieval to be deemed
12 acceptable, where "AND" and "OR" are Boolean operators. If one
13 of these two conditions is not satisfied, module 7 determines
14 that the retrieval is suspicious.
15

17 Alternative to a preselected integral margin such as the
18 margin of 1 on either side of the observed numbers of rows and
19 columns illustrated above, any statistical technique may be used
20 to generate the rules of table 6 from the corresponding retrieval
21 information. For example, the margin on the positive side of the
22 number of observations may be a preselected percent of the
23 observed value, or a preselected number of standard deviations.
24 The margin on the lower side of the observed value may be the
25 same or a different percent of the observed value, or the same or
26 a different number of standard deviations. Other statistical
27
28

1 techniques will be readily attainable by those of ordinary skill
2 in the art.

3 Figure 4 illustrates two optional steps, steps 41 and 42,
4 that constitute the system administrator 10 phase. At step 41,
5 suspicious activity that is observed during the optional training
6 phase is reported to system administrator 10. For example, if
7 the retrieval of a certain combination of rows and columns during
8 the training phase is observed to occur fewer than a preselected
9 threshold number of times, such activity can be flagged to the
10 system administrator 10 as being suspicious. In the above
11 example, suppose that, in addition to five rows and three columns
12 being retrieved 963 times and seven rows and two columns being
13 retrieved 51 times, one row and 100 columns were retrieved one
14 time. This might indicate that the requesting user 3 is
15 attempting to retrieve too much information in a single command
16 5, and this activity is reported to the system administrator 10
17 at step 41 as being suspicious.

18 Similarly, one could incorporate within parameters 17 a
19 maximum number of rows allowed to be retrieved (possibly for a
20 given field/table or set of fields/tables). Let us assume that
21 this maximum number of rows is 20. Then if a particular training
22 phase retrieval attempts to retrieve 21 or more rows, such a
23 retrieval is deemed to be suspicious and is likewise reported to
24 system administrator 10 at step 41. System administrator 10 can
25

1 then remove from the set of acceptable retrievals within table 6
2 such suspicious retrievals.

3 At step 42, system administrator 10 can force feed rules
4 into table 6. Step 42 can be performed in lieu of or in addition
5 to the training phase. For example, one of the rules provided by
6 the system administrator 10 could be: "no more than 100 rows
7 from CREDIT CARD table are acceptable" or "no more than 100 rows
8 in any one minute from CREDIT CARD table are acceptable".
9

10 Rules can also be entirely statistical, such as:

11 "If the number of rows retrieved by a single user to the
12 CREDIT card field exceeds the historical average for the user's
13 group by more than 2 standard deviations, then generate an
14 alert."
15

16 The above description is included to illustrate the
17 operation of the preferred embodiments and is not meant to limit
18 the scope of the invention. The scope of the invention is to be
19 limited only by the following claims. From the above discussion,
20 many variations will be apparent to one skilled in the art that
21 would yet be encompassed by the spirit and scope of the present
22 invention. For example, instead of training the system 19 on the
23 number of columns overall, one could single out certain columns
24 (or combinations of columns) of interest within database 1 and
25 train on that basis, e.g., one could train on the SOCIAL SECURITY
26
27
28

1 NUMBER column within the PAYROLL table, and/or the CREDIT CARD

2 NUMBER column within the CREDIT INFORMATION table.

3 What is claimed is:

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28